Stochastik für die Informatik

Wintersemester 2020/21

Anton Wakolbinger

Stofl-Webseite:

https://www.math.uni-frankfurt.de/

$\sim$ismi/wakolbinger/teaching/StofI2021/

`https://www.math.uni-frankfurt.de/~ismi/wakolbinge`

Stofl-Webseite:

https://www.math.uni-frankfurt.de/

∼ismi/wakolbinger/teaching/StofI2021/

`https://www.math.uni-frankfurt.de/~ismi/wakolbinge`

Oder:

google: “Wakolbinger”

→ Goethe-Universität – Homepage A. Wakolbinger

→ Veranstaltungen WS 2020/21

Dort: Link zum

Moodle-Kurs “Stochastik für die Informatik”

Anmeldung zu einer **Übungsgruppe:**

elektronisch ($\leftarrow$ Stofl - Webseite)

bis Donnerstag 05.11.2020, 24 Uhr

nach dem first come - first serve Prinzip

Gruppenzeiten:

Di 14-16, Di 16-18,

Mi 12-14, 12-14

Do 12-14 , Do 12-14, Do 14-16

Fr 8-10, Fr 10-12, Fr 14-16

Seit gestern (02.11.) sind alle 10 Gruppen (à 40) voll.

Wenn Sie sich noch zu den Übungen anmelden wollen, dann schreiben Sie bitte eine E-Mail an den Übungskoordinator: Herrn Florin Boenkost, boenkost@math.uni-frankfurt.de mit der Angabe von Ihnen möglichen Terminen.

Und bitte halten Sie Ihre Anmeldung nur dann aufrecht, wenn Sie wirklich an den Übungen teilnehmen wollen.

jeweils spätestens Dienstags:

Neues Übungsblatt im Moodle-Kurs

Tipps zu den Übungsaufgaben:

in den Tutorien

(und zu Blatt 1 auch schon in der Vorlesung diesen Freitag)

Termin für die elektronische Abgabe

der schriftlichen Lösungen der “S-Aufgaben”:

Freitags (10 Tage nach Ausgabe des Blattes)

Leitfaden zur elektronischen Abgabe: siehe Moodle-Kurs.

In der Woche nach der Abgabe

werden die Lösungen in den Tutorien besprochen.

Abschlussklausur: Dienstag, 16 Februar 2021

Zweitklausur: Donnerstag 08. April 2021.

Beide Termine sind derzeit noch unter Vorbehalt.

Abschlussklausur: Dienstag, 16 Februar 2021

Zweitklausur: Donnerstag 08. April 2021.

Beide Termine sind derzeit noch unter Vorbehalt.

In der Klausur können 100 Klausurpunkte erreicht werden.

Die Note errechnet sich aus

der Summe der Anzahl der erreichten Klausurpunkte

plus der Anzahl der im Tutorium erreichten Bonuspunkte.

Abschlussklausur: Dienstag, 16 Februar 2021

Zweitklausur: Donnerstag 08. April 2021.

Beide Termine sind derzeit noch unter Vorbehalt.

In der Klausur können 100 Klausurpunkte erreicht werden.

Die Note errechnet sich aus

der Summe der Anzahl der erreichten Klausurpunkte

plus der Anzahl der im Tutorium erreichten Bonuspunkte.

Beträgt diese Summe mindestens 45,
gilt die Abschlussprüfung über die Veranstaltung
als bestanden.

Bonuspunkte (maximal 12):

durch aktive Beteiligung in den Tutorien.

Bonuspunkte bekommt man nur, wenn man
mindestes zweimal im Semester
Lösungen von Übungsaufgaben (oder Teile davon)
im Tutorium vorstellt,

und grundsätzlich nur für die Aufgaben, bei deren
Lösungsbesprechung man im Tutorium anwesend ist.

Lehrbuch:

Götz Kersting, Anton Wakolbinger

Elementare Stochastik, Birkhäuser, 2. Aufl. 2010,

Preis: 19,99 EUR

Semesterausleihe möglich aus der

Bibliothek des Mathematischen Seminars,

Robert-Mayer-Str. 8, 4. Stock

in der UB als E-Book vorhanden

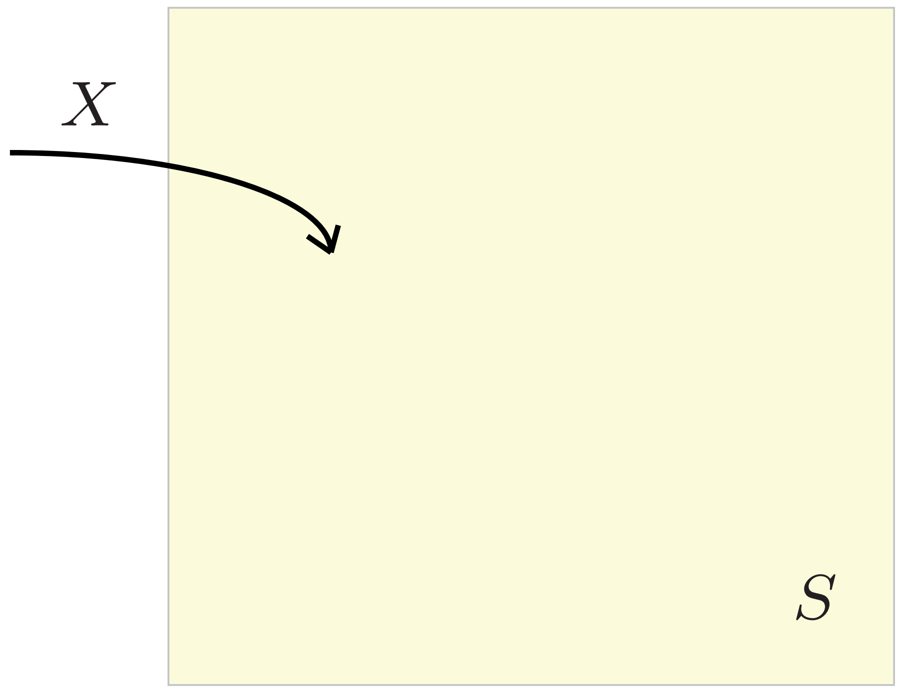
Vorlesung 1a

Zufallsvariable und Ereignisse, Wahrscheinlichkeiten und Verteilungen

Ein erster Blick

1. Rein zufällige Wahl aus einer endlichen Menge

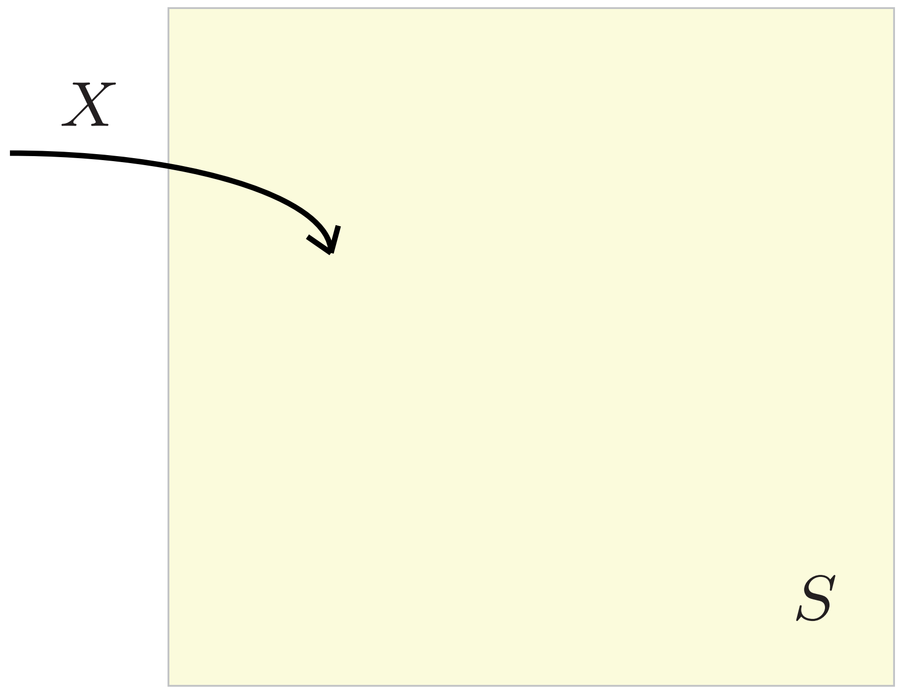
Prototypisches Beispiel einer Zufallsvariablen:



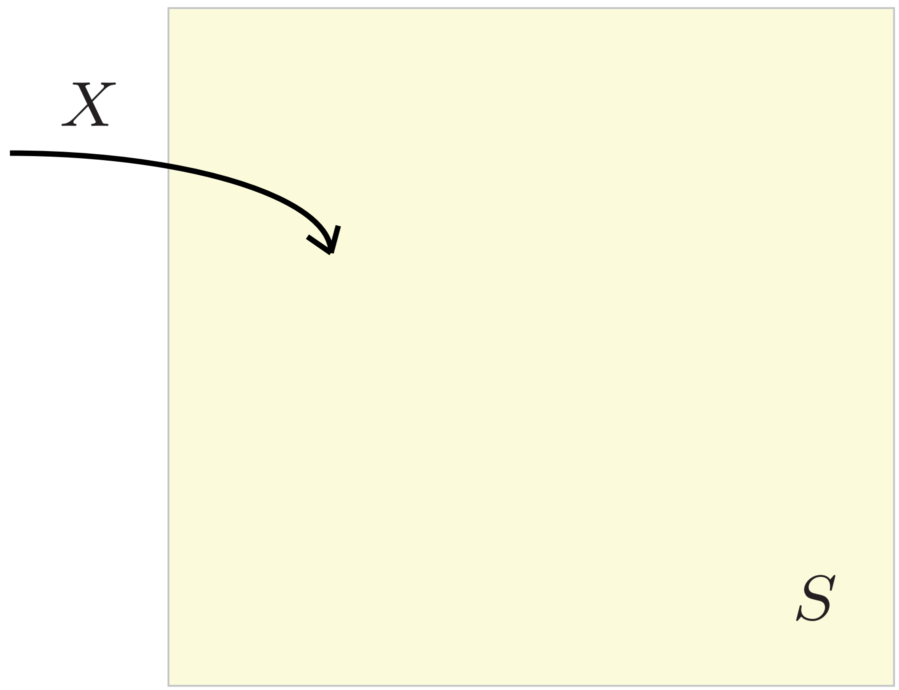
rein zufällige Wahl eines Punktes aus einer Fläche

(z. B. aus dem beigen Quadrat S)

Prototypisches Beispiel einer Zufallsvariablen:

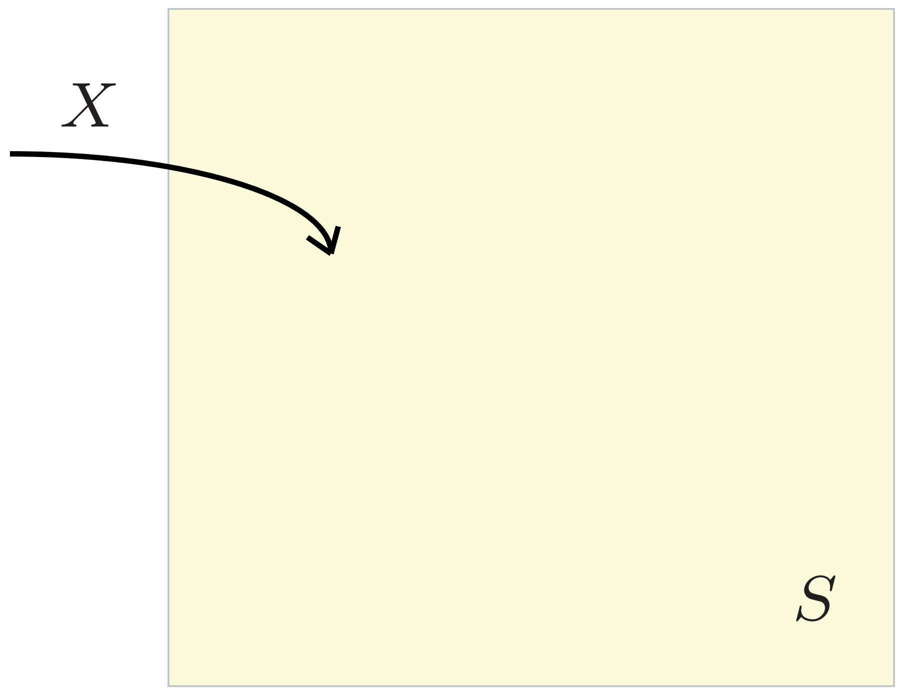


Was heißt “rein zufällige Wahl” eines Punktes aus S?

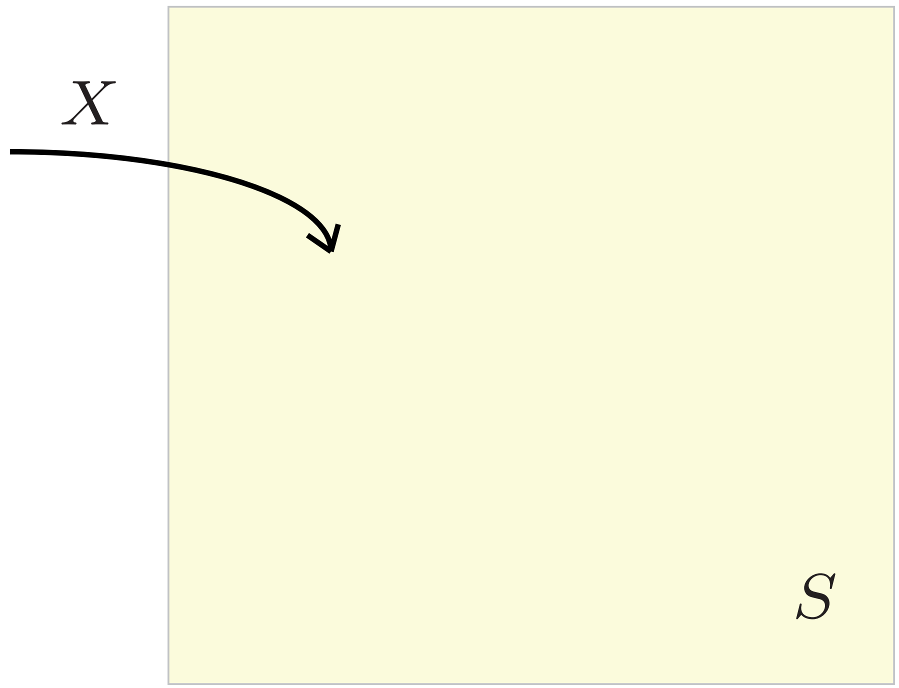


Fürs Erste sind endlich viele Pixel leichter vorstellbar
als ein Kontinuum aus unendlich vielen Punkten.

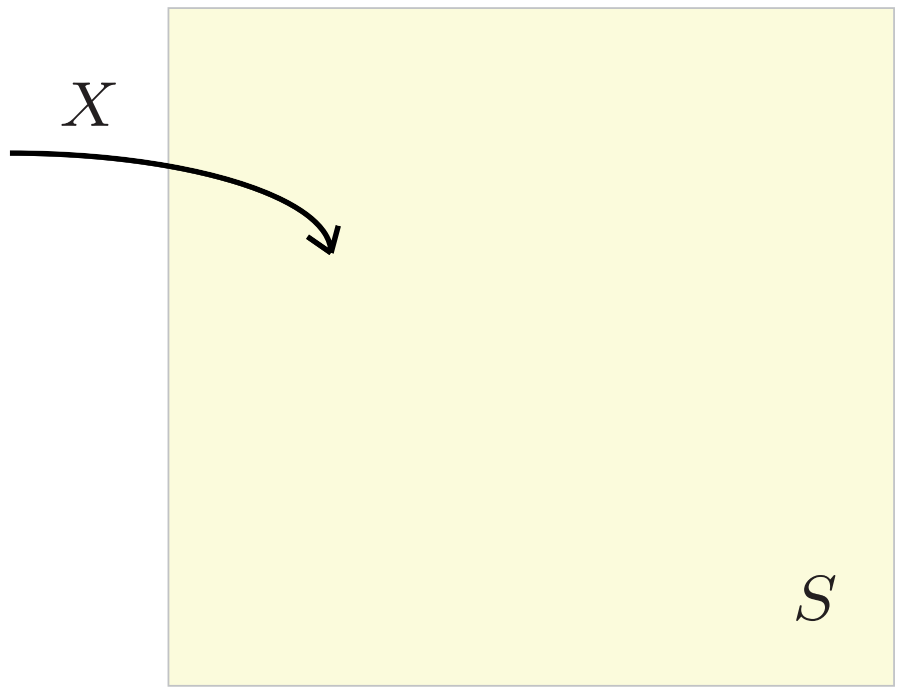
Stellen wir uns vor, S besteht aus 1000×1000 Pixeln



Stellen wir uns vor, S besteht aus g Pixeln, mit $g \in \mathbb{N}$

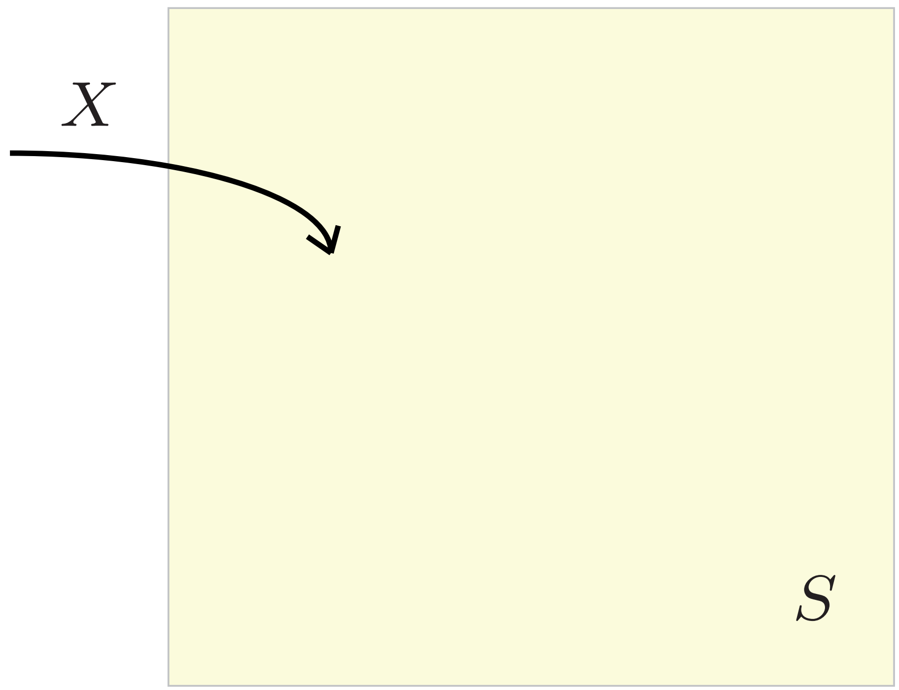


“Rein zufällige Wahl aus S” soll heißen:



alle Pixel in S haben dieselbe Chance,

zum Zug zu kommen.



Man spricht dann von einer

uniform auf S verteilten Zufallsvariablen X

Analogie zum **fairen Würfeln**:

Die Menge der möglichen Ausgänge ist hier

$$S = \{1, 2, 3, 4, 5, 6\}.$$

Auch hier wird der zufällige Ausgang beschrieben durch eine uniform auf S verteilte Zufallsvariable X.

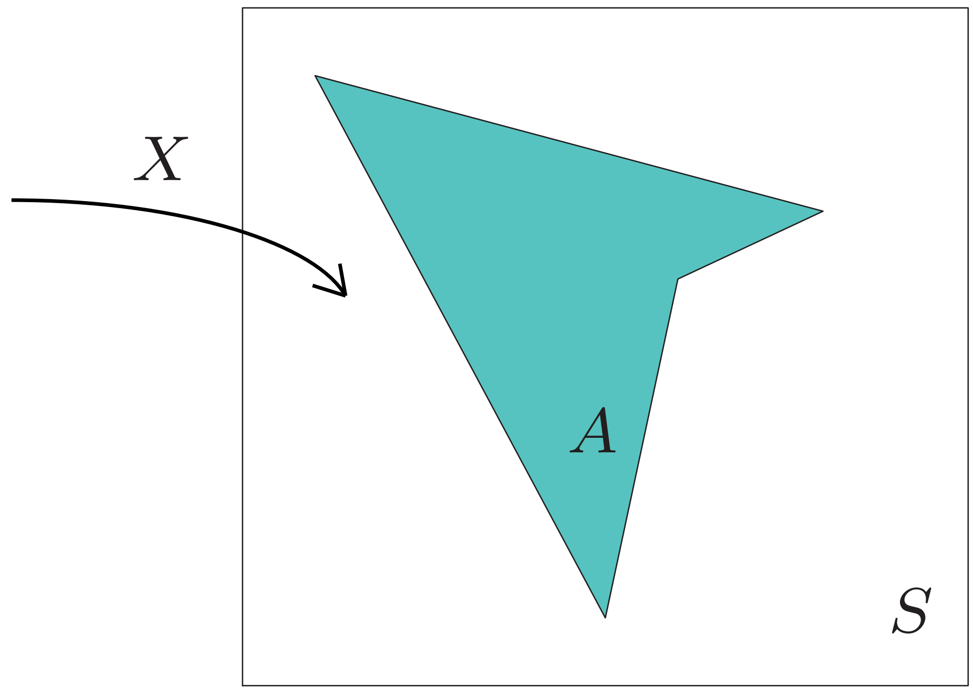
Wir halten fest:

Zufallsvariable mit Werten in einer endlichen Menge S
entsprechen einer zufälligen Wahl aus S.

Auf S **uniform** verteilte Zufallsvariable
entsprechen einer **rein** zufälligen Wahl aus S,
bei der alle Elemente von S die gleiche Chance haben,
zum Zug zu kommen.

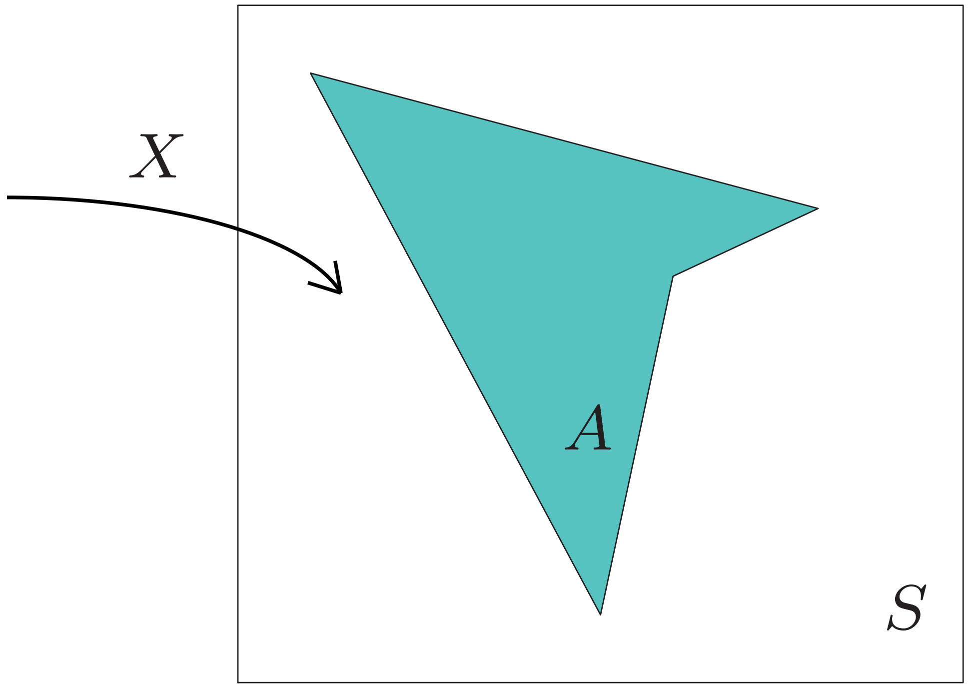
2. Ereignisse

Betrachten wir wieder unser Quadrat S



diesmal zusammen mit einer bestimmten Teilmenge A von S

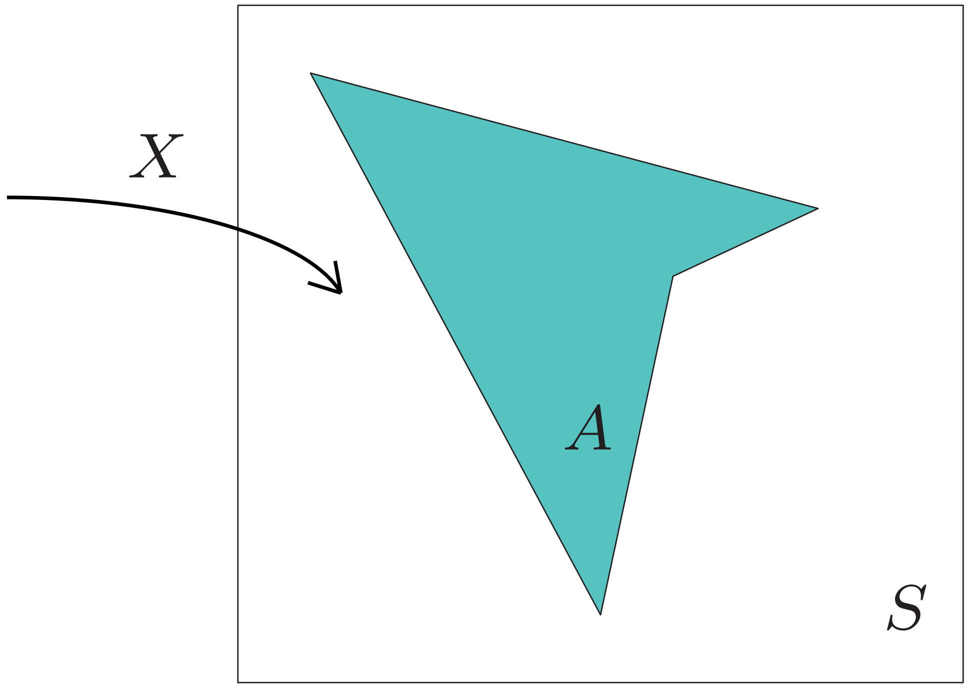
Bei der Wahl eines Pixels aus S



kann die Wahl auf A fallen

– oder auch nicht.

Bei der Wahl eines Pixels aus S



kann das Ereignis “X fällt in A” eintreten

– oder auch nicht.

Das Ereignis “X fällt in A”

notiert man als

$$\{X \in A\}.$$

Die Menge aller Ereignisse $\{X \in A\}, \;\; A \subset S$,

nennt man auch die

“von X erzeugte Kollektion von Ereignissen”.

Die Menge aller Ereignisse $\{X \in A\}, \;\; A \subset S$,

nennt man auch die

"von X erzeugte Kollektion von Ereignissen".

Ereignisse kann man (aussagen-)logisch verknüpfen, z.B. gilt:

$\{X \in A\}$ und $\{X \in B\} = \{X \in A \cap B\}$.

Mehr dazu später!

3. Wahrscheinlichkeiten

Wie wahrscheinlich ist es,

dass bei einer rein zufälligen Wahl eines Pixels aus S

die Wahl auf A fällt?

Wie wahrscheinlich ist es,

dass bei einer rein zufälligen Wahl eines Pixels aus S

die Wahl auf A fällt?

Wie groß ist die Wahrscheinlichkeit

des Ereignisses $\{X \in A\}$?

Wie wahrscheinlich ist es,

dass bei einer rein zufälligen Wahl eines Pixels aus S

die Wahl auf A fällt?

Wie groß ist die Wahrscheinlichkeit

des Ereignisses $\{X \in A\}$?

.... eine Zahl zwischen 0 $(= 0\%)$ und 1 $(= 100\%)$

Merke: Das Ereignis $\{X \in S\}$ hat Wahrscheinlichkeit 1.

Für die Wahrscheinlichkeit des Ereignisses $\{X \in A\}$

schreiben wir

$$\mathbf{P}(\{X \in A\})$$

oder kurz

$$\mathbf{P}(X \in A).$$

Für die Wahrscheinlichkeit des Ereignisses $\{X \in A\}$

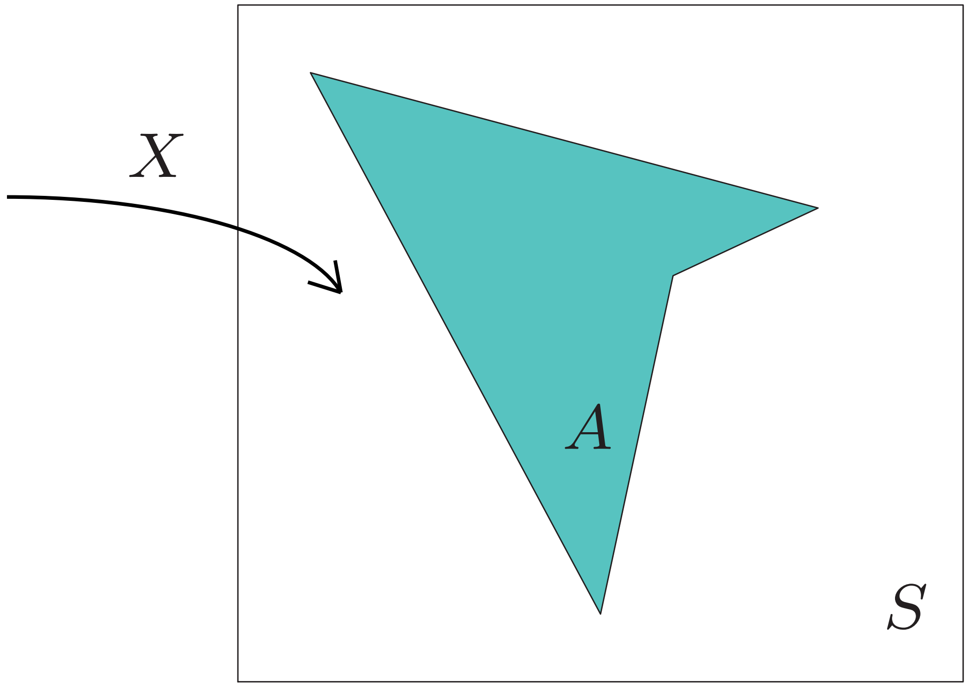
schreiben wir

$$\mathbf{P}(\{X \in A\})$$

oder kurz

$$\mathbf{P}(X \in A).$$

P steht für **probabilitas** (= Wahrscheinlichkeit)



Bei der rein zufälligen Wahl eines Pixels aus S,

beschrieben durch die Zufallsvariable X, ist

die Wahrscheinlichkeit des Ereignisses $\{X \in A\}$

proportional zur Anzahl der Pixel in A.

Zusammengefasst:

X ist *rein zufälliger Pixel aus dem Quadrat* S

bedeutet:

Für jede Teilmenge A von S ist

$$\mathbf{P}(\{X \in A\}) = \frac{\text{Anzahl der Pixel in } A}{\text{Anzahl der Pixel in } S}.$$

Lies und merke:

die Wahrscheinlichkeit, dass X in A fällt

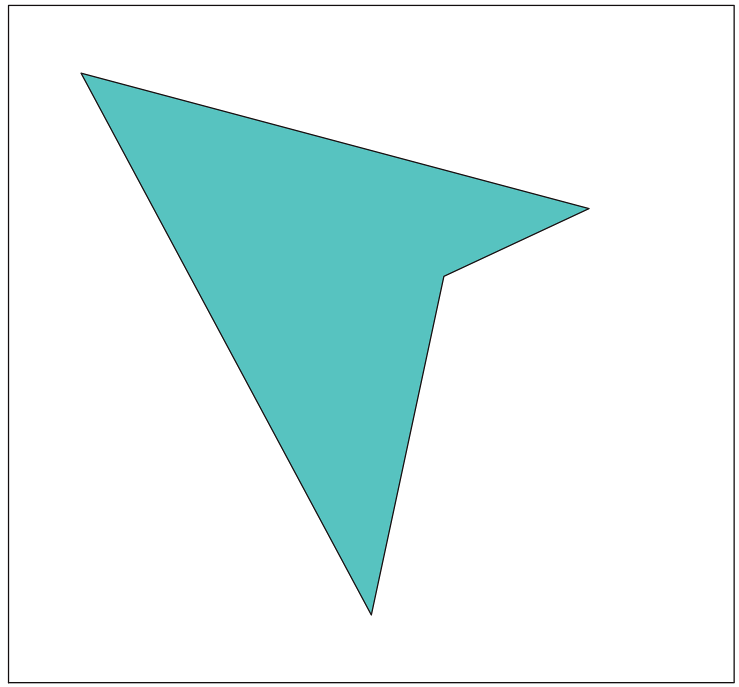
ist der Anteil der Menge A an der Menge S.

4. Schätzung eines Flächenanteils

Eine Anwendung der “rein zufälligen Wahl”:

Monte-Carlo Schätzung eines Flächenanteils.

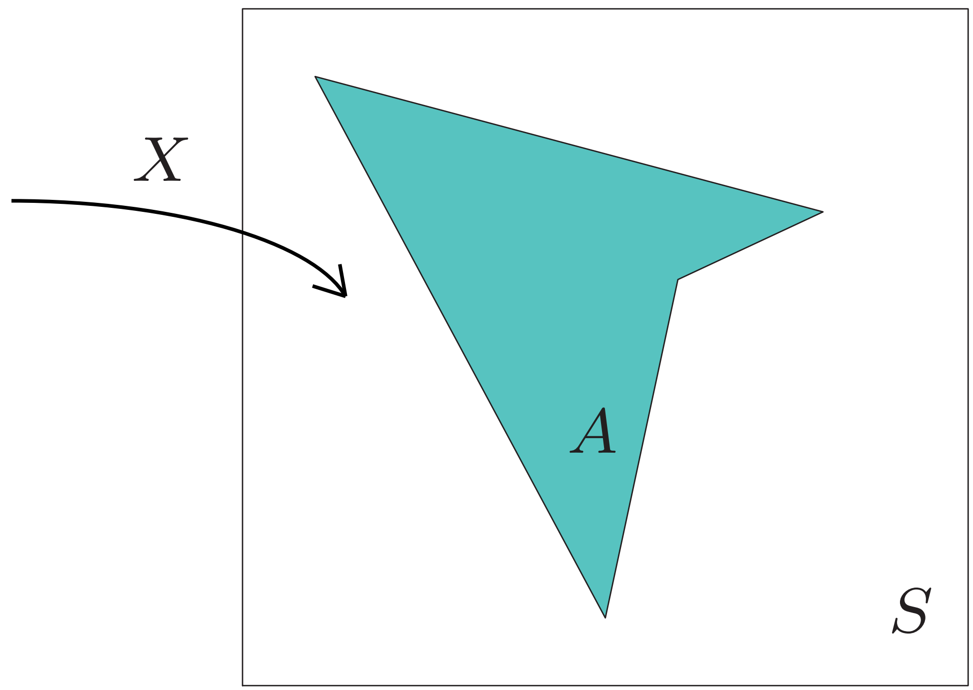
Wir fragen:



Wie groß ist der Anteil der blauen Fläche

an der Fläche des Quadrats?

und übersetzen in die Sprache der Stochastik:



Wie wahrscheinlich ist es, dass die

rein zufällige Wahl eines Pixels aus dem Quadrat S

in die blaue Fläche trifft?

Wie wir bald sehen werden

(und wie auch intuitiv klar ist)

gibt es einen engen Zusammenhang zwischen

Wahrscheinlichkeiten und Trefferquoten.

Angenommen wir haben ein Werkzeug, mit dem man einen **rein zufälligen Pixel** aus dem Quadrat wählen kann

– und das nicht nur einmal, sondern “immer wieder neu”.

Wir bekommen dann mit unserem Werkzeug

nicht nur *einen einzigen* rein zufälligen Pixel,

sondern sogar beliebig viele,

genauer:

eine *rein zufällige Folge* $(X_1, X_2, ...)$

von Pixeln in S.

Sei $A \subset S$.

$$Z_i := \mathbf{1}_A(X_i), \quad i = 1, 2, \ldots$$

zählt, ob X_i in A fällt.

Sei $A \subset S$.

$$Z_i := \mathbf{1}_A(X_i), \quad i = 1, 2, \ldots$$

zählt, ob X_i in A fällt.

Dabei ist

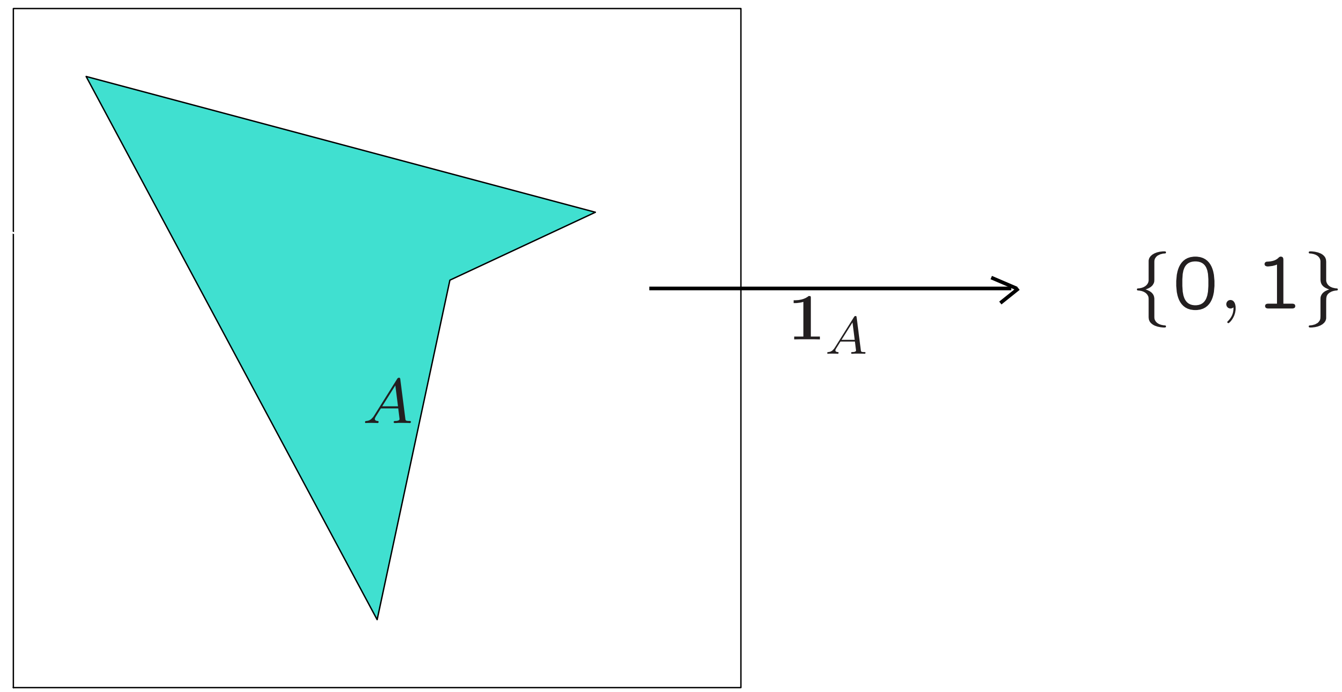
$$\mathbf{1}_A(x) = \begin{cases} 1 \text{ für } x \in A \\ 0 \text{ für } x \in S \setminus A \end{cases}$$

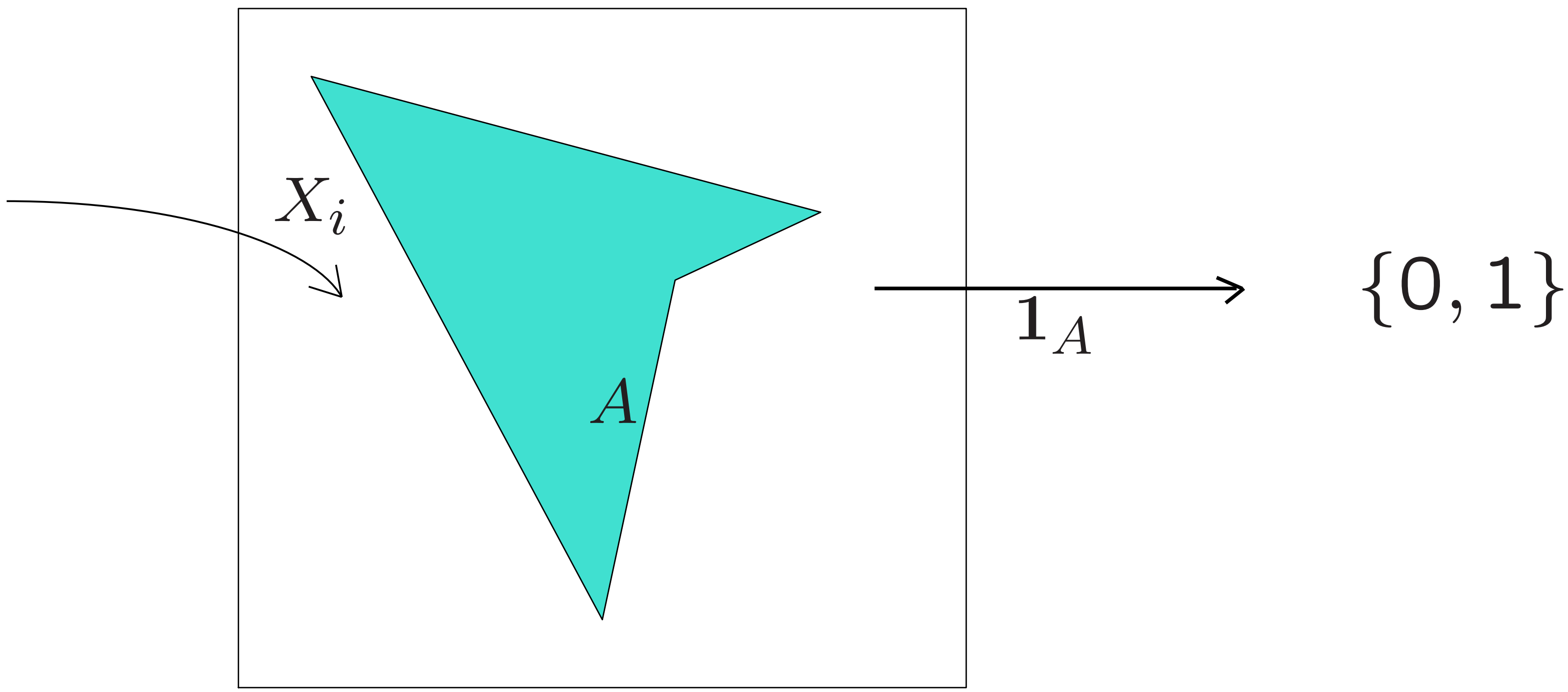
die Indikatorfunktion der Menge A.

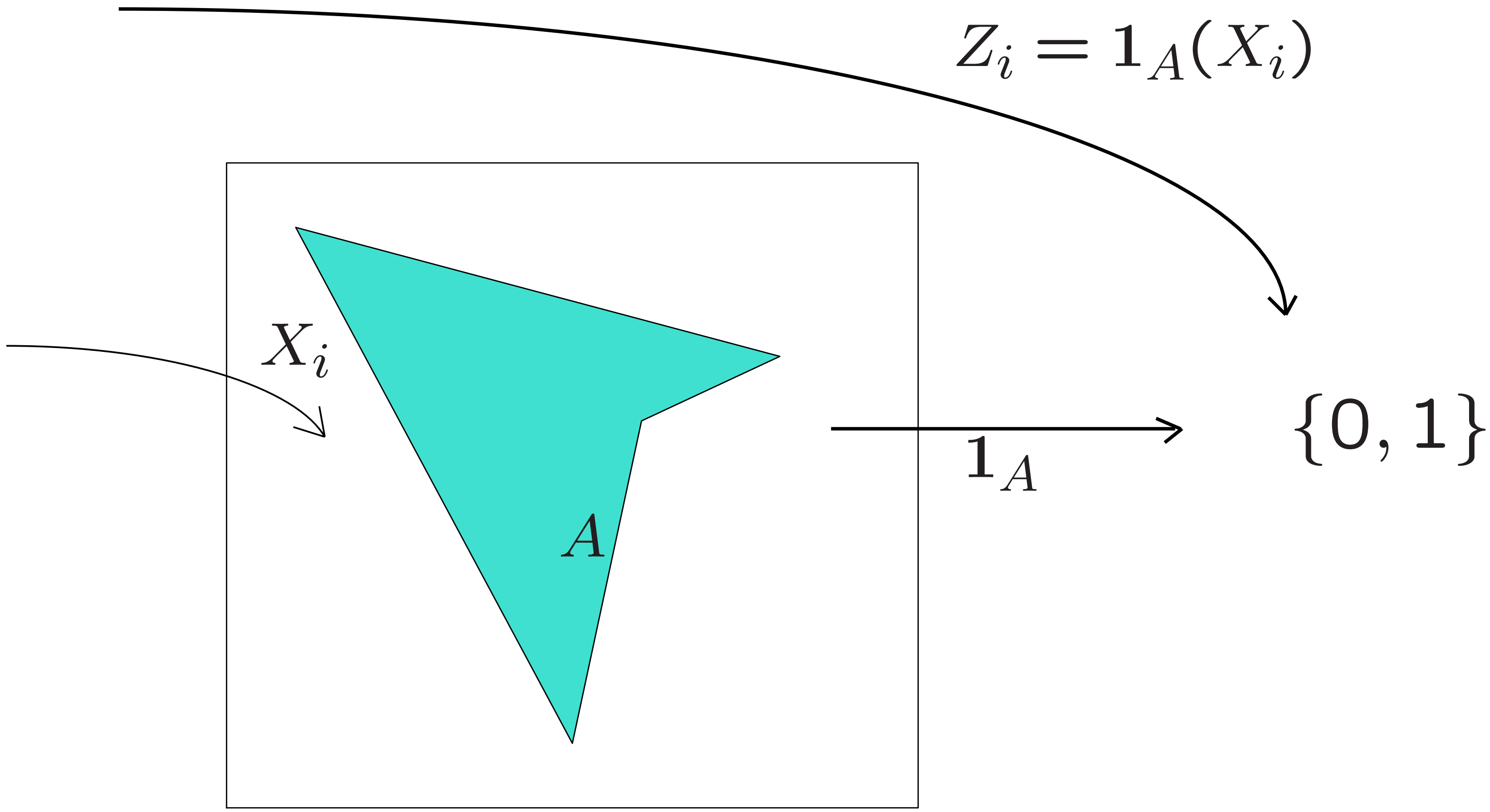
Mit dieser Abbildung von S nach $\{0, 1\}$

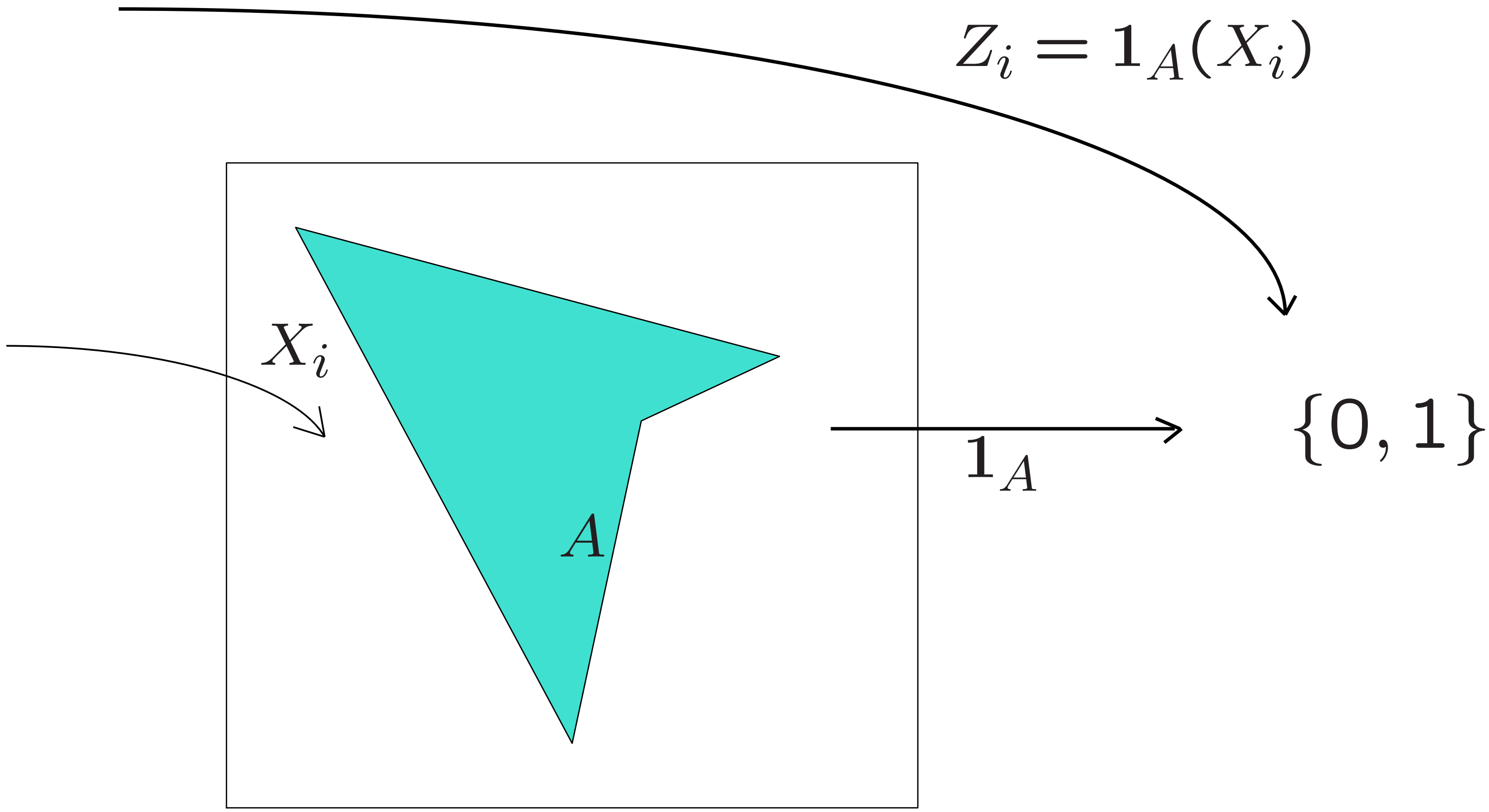
wird die S-wertige Zufallsvariable X_i

zur $\{0, 1\}$-wertigen Zufallsvariablen Z_i “verarbeitet”.


A
$\mathbf{1}_A$
$\{0, 1\}$

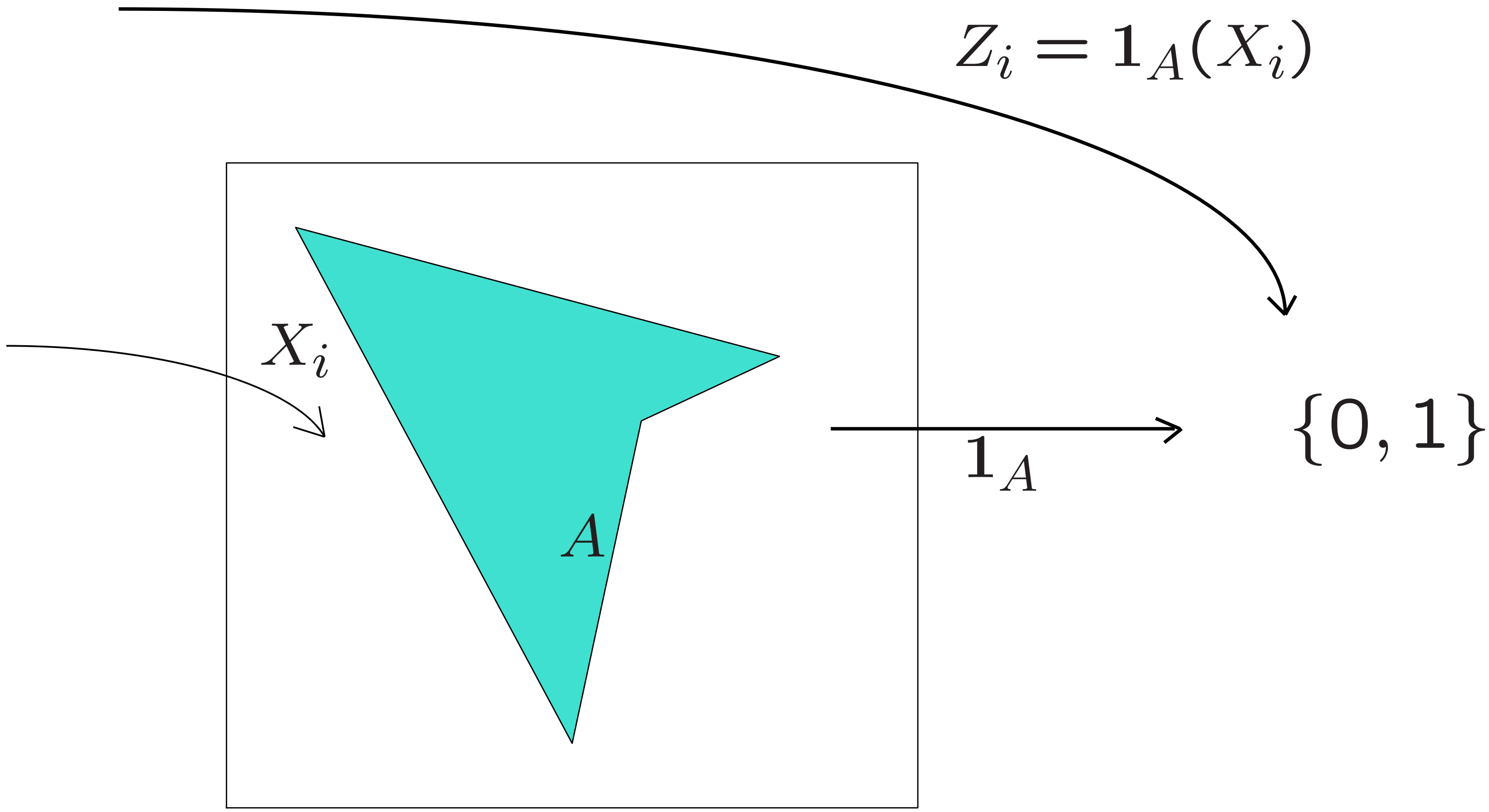

X_i
A
$\mathbf{1}_A$
$\{0, 1\}$


$Z_i = \mathbf{1}_A(X_i)$
X_i
A
$\mathbf{1}_A$
$\{0, 1\}$

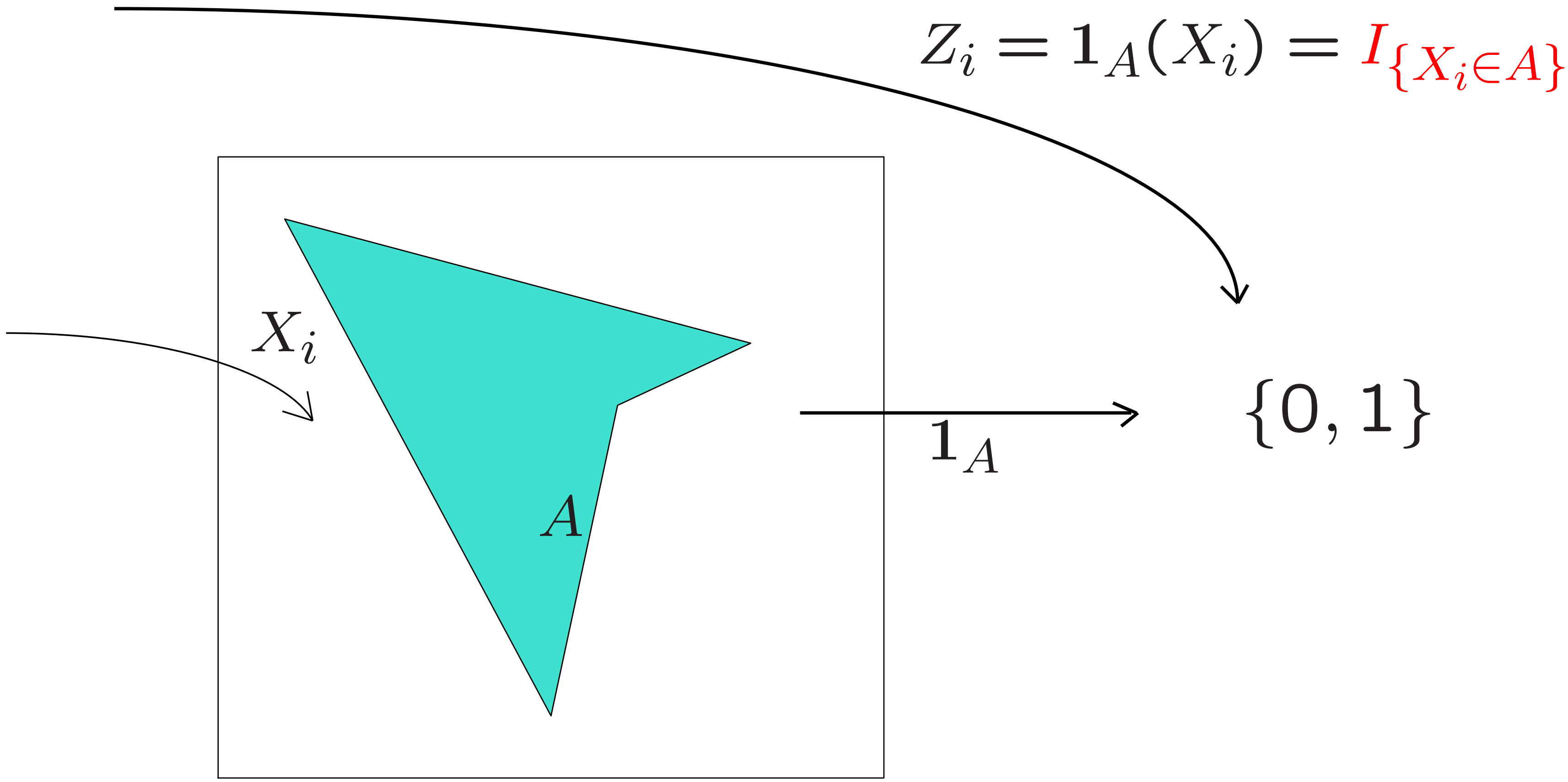


$Z_i = \mathbf{1}_A(X_i)$ ist der *der Indikator* *("der Zähler")*

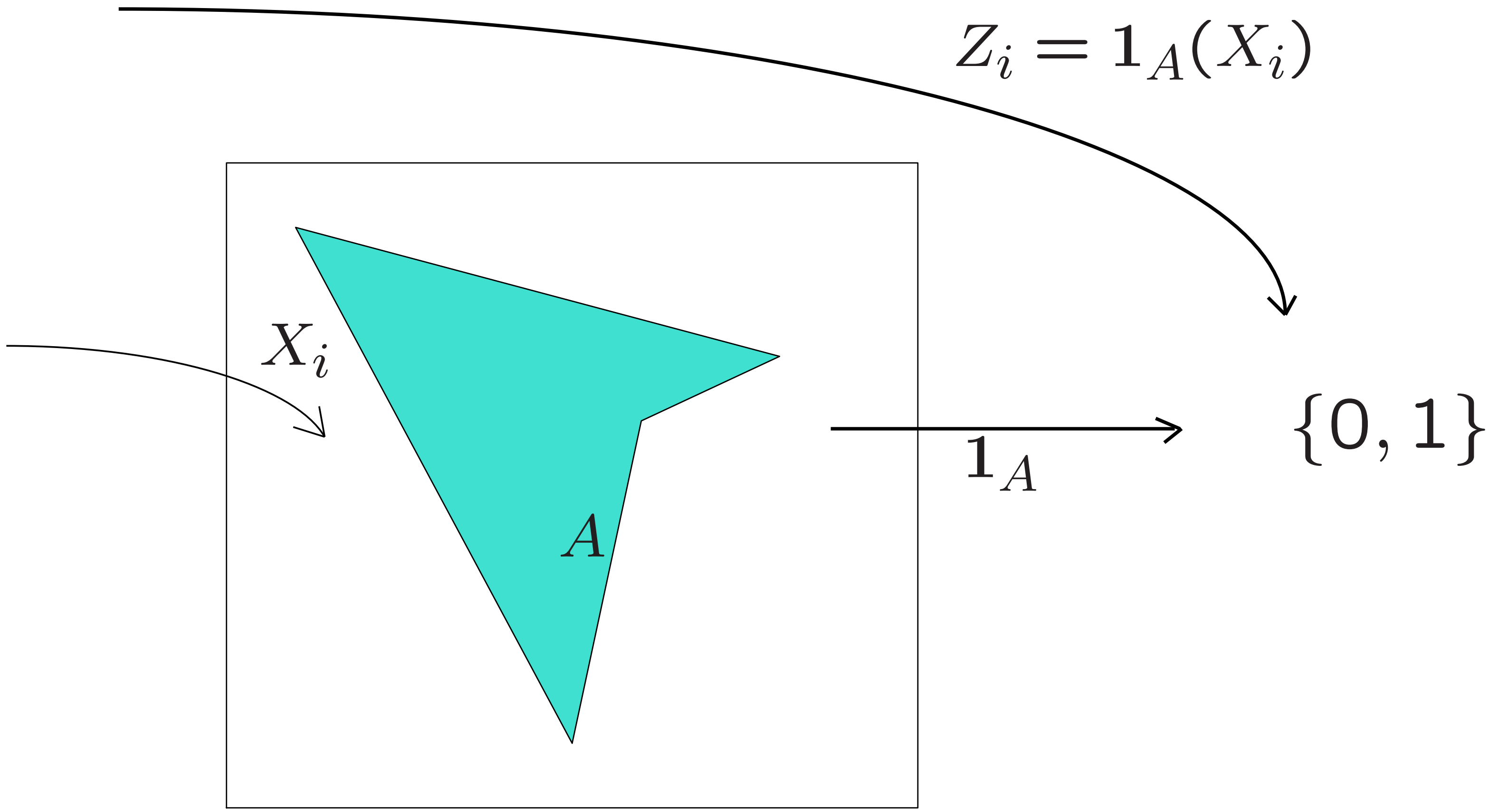
des Ereignisses "X_i fällt in A".



Eine alternative Schreibweise für $\mathbf{1}_A(X_i)$ ist $I_{\{X_i \in A\}}$



Eine alternative Schreibweise für $\mathbf{1}_A(X_i)$ ist $I_{\{X_i \in A\}}$



$\{X_i \in A\}$ und $\{Z_i = 1\}$ sind zwei Schreibweisen für

dasselbe Ereignis "X_i fällt in A"

5. Die Verteilung der zufälligen Trefferquote

Illustriert durch ein R-Programm

werden $n = 100$ Punkte

“rein zufällig” aus dem Quadrat gewählt.

Illustriert durch ein R-Programm

werden $n = 100$ Punkte

"rein zufällig" aus dem Quadrat gewählt.

$$Z_i = \mathbf{1}_A(X_i), \;\; i = 1, 2, \ldots$$

Illustriert durch ein R-Programm

werden $n = 100$ Punkte

“rein zufällig” aus dem Quadrat gewählt.

$$Z_i = \mathbf{1}_A(X_i), \;\; i = 1, 2, \ldots$$

Die zufällige Zahl

$$M := \frac{1}{100}(Z_1 + \cdots + Z_{100})$$

(die “Trefferquote”)

ist ein *Schätzer* für die Wahrscheinlichkeit

$$p := \mathbf{P}(X \in A)$$

(und damit für den gefragten Flächenanteil).

Ein Ergebnis ("eine Realisierung") von $(X_1, \ldots, X_{100})$

liefert eine Realisierung von $(Z_1, \ldots, Z_{100})$

und damit eine Realisierung von M

(einen Schätzwert für p).

Wie "zuverlässig" ist dieser Schätzwert?

$$M := \frac{1}{100}(Z_1 + \cdots + Z_{100})$$

Es sei an dieser Stelle verraten:

Der Anteil der blauen Fläche am Quadrat

(den man in der Realität ja nicht kennt) ist in unserem Beispiel

$$p = 0.195$$

$$M := \frac{1}{100}(Z_1 + \cdots + Z_{100})$$

Es sei an dieser Stelle verraten:

Der Anteil der blauen Fläche am Quadrat

(den man in der Realität ja nicht kennt) ist in unserem Beispiel

$$p = 0.195$$

Damit hat M gar keine Chance, exakt auf p zu fallen, denn

der *Wertebereich* von M

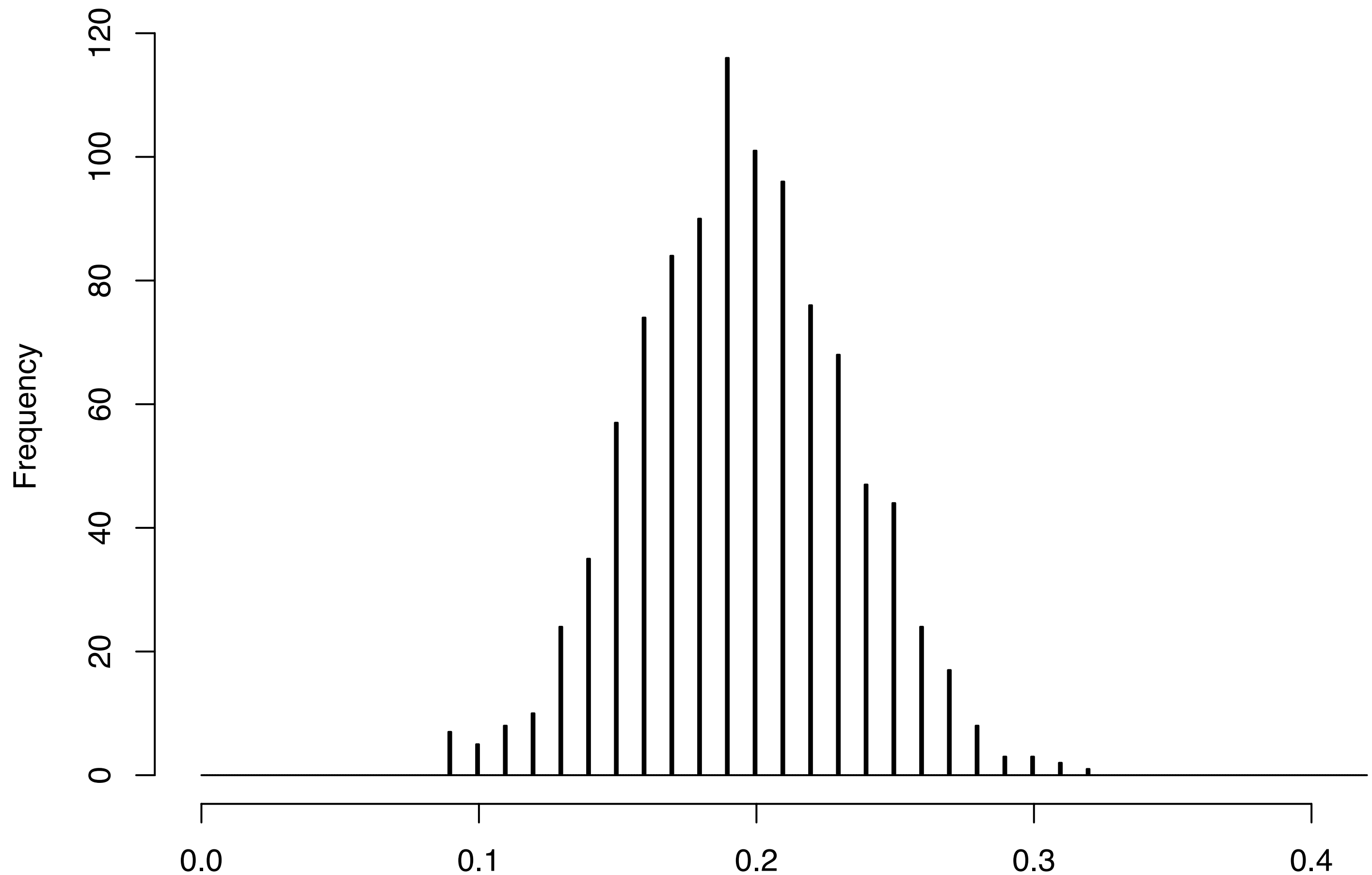
(d.h. die Menge der möglichen Ausgänge) ist

$$S' := \left\{\frac{0}{100}, \frac{1}{100}, \ldots, \frac{99}{100}, \frac{100}{100}\right\}$$

Wie zuverlässig ist M als Schätzer für p?

Davon machen wir uns ein Bild, indem wir viele (z.B. 1000) “unabhängige Kopien” von M erzeugten und in einem *Histogramm* darstellen, wie oft welche Ausgänge realisert wurden.

Verteilung von M (Näherung aus 1000 Wiederholungen):



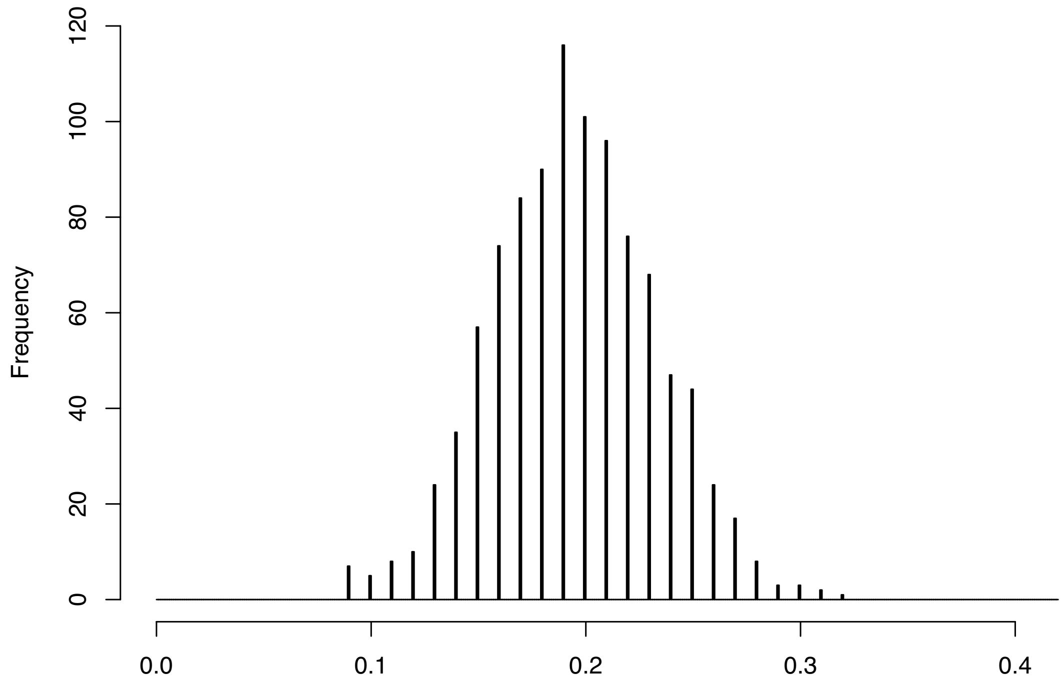
Wie zuverlässig ist M als Schätzer für p?
Davon machen wir uns ein Bild, indem wir viele (z.B. 1000) “unabhängige Kopien” von M erzeugten und in einem *Histogramm* darstellen, wie oft welche Ausgänge realisert wurden.

So bekommen wir eine näherungsweise Darstellung der Verteilung von M.

Die Verteilung von M ist bestimmt durch ihre Gewichte

$$\rho(b) := \mathbf{P}(M = b), \quad b \in S'.$$

Verteilung von M (Näherung aus 1000 Wiederholungen):



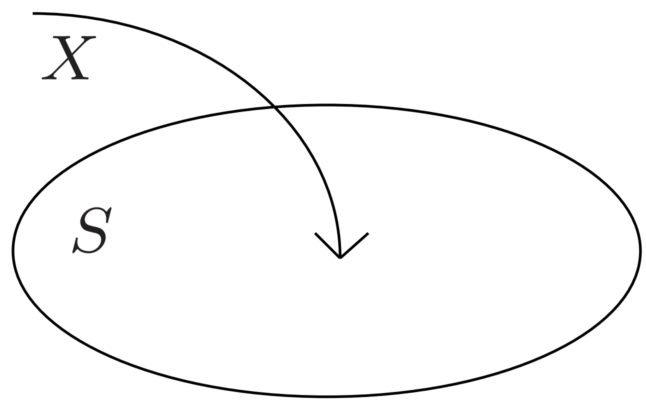
Die 101 möglichen Ausgänge von M

sind (bei weitem) nicht gleich wahrscheinlich:

die Verteilung von M “ist um p konzentriert”.

6. Zusammenfassung der wichtigsten Begriffe der ersten Stunde

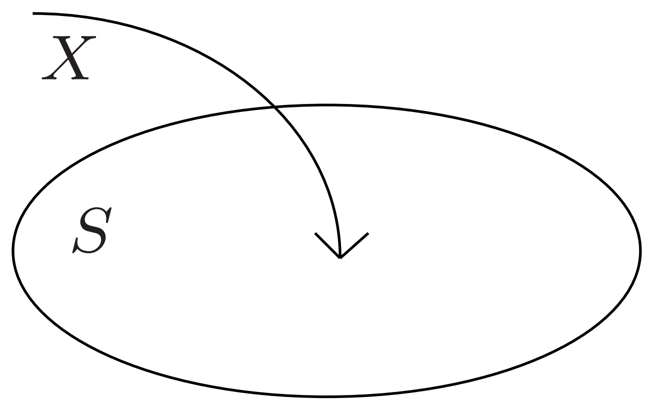
Ein Logo der Elementaren Stochastik:



X ... zufällige Wahl eines Elements aus S

S ... Menge von möglichen Ausgängen

Ein Logo der Elementaren Stochastik:

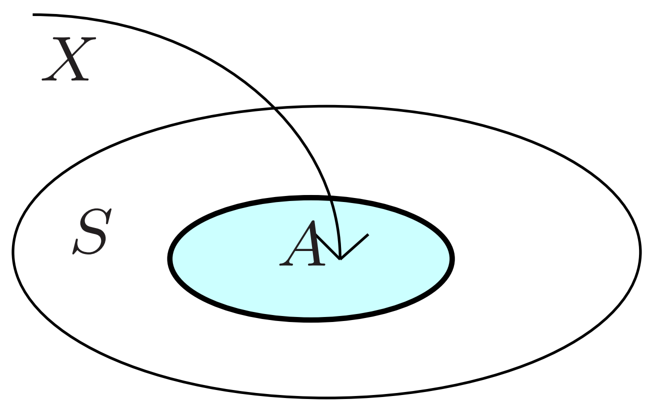


X ... **Zufallsvariable**

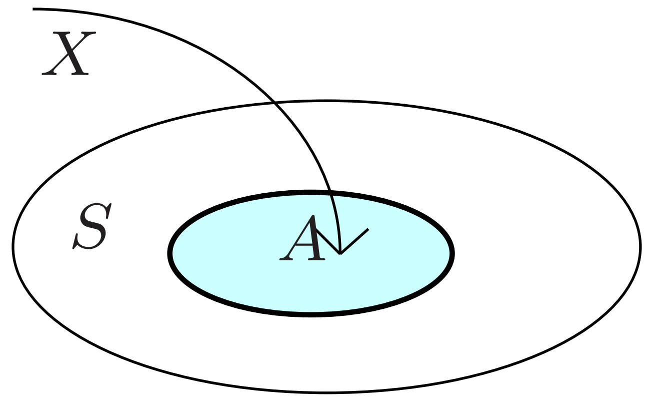
mit *Zielbereich (Wertebereich)* S

Wir interessieren uns für die *Wahrscheinlichkeit*

des *Ereignisses* "X fällt in A"



Dabei ist A eine bestimmte Teilmenge von S.



Ereignisse werden (wie Mengen)

in geschweiften Klammern notiert:

$$\{X \in A\}$$

Lies:

“X fällt in A”.

“X ist *rein zufällig*”

heißt im Fall einer *endlichen* Menge S:

alle Elemente von S haben die gleiche W’keit

gewählt zu werden.

“X ist *rein zufällig*”

heißt im Fall einer *endlichen* Menge S:

alle Elemente von S haben die gleiche W’keit

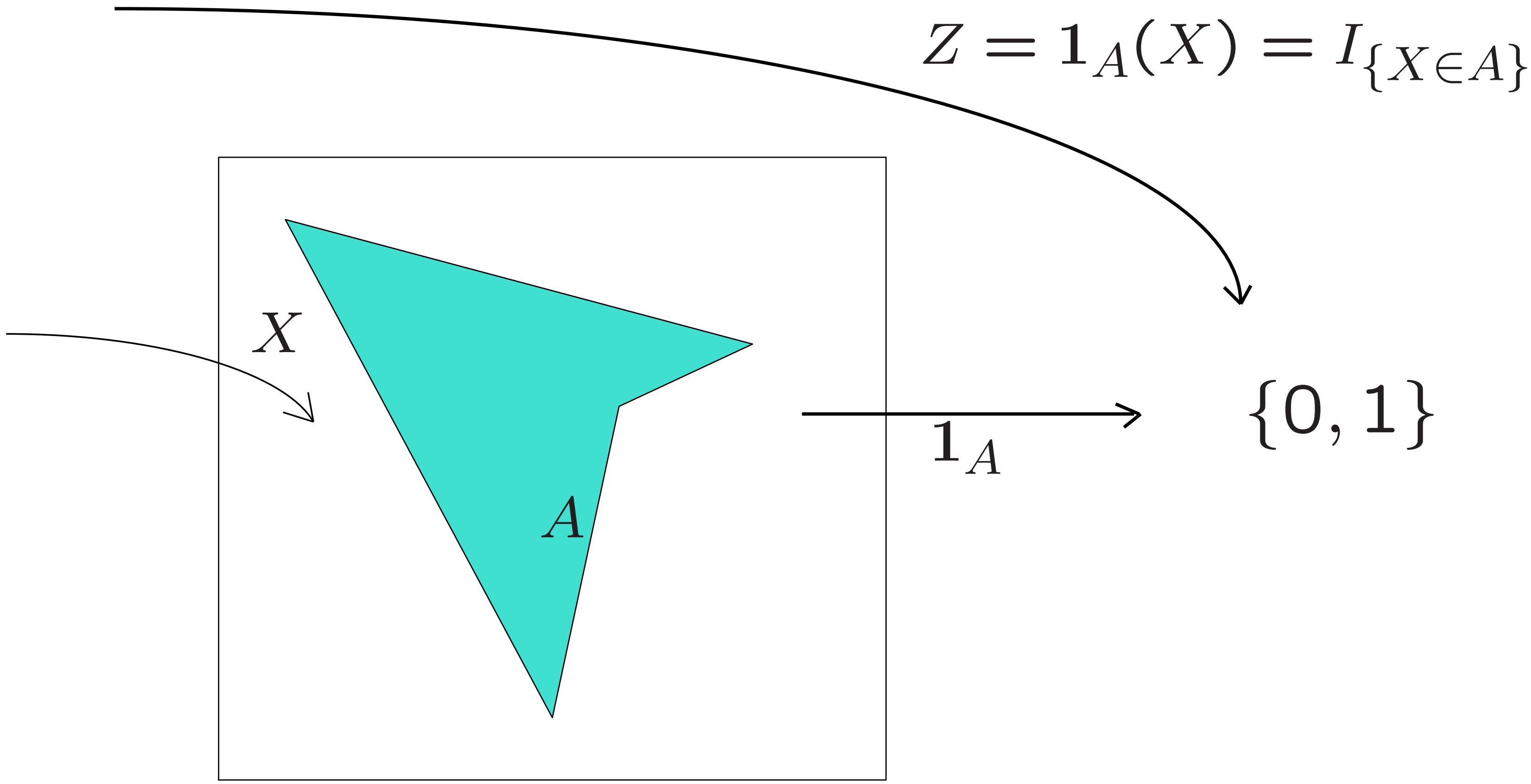
gewählt zu werden.

Dann ist die Wahrscheinlichkeit des Ereignisses “X fällt in A”:

$$\mathbf{P}(\{X \in A\}) = \frac{\# A}{\# S}.$$

Statt $\mathbf{P}(\{X \in A\})$ scheiben wir kurz:

$$\mathbf{P}(X \in A).$$



Die Ereignisse $\{X \in A\}$ und $\{Z = 1\}$ stimmen überein!

Ein Ausblick ins Kontinuum:

“X ist *rein zufällige Wahl aus S*”

heißt im Fall einer **kontinuierlichen** Menge $S \subset \mathbb{R}^d$:

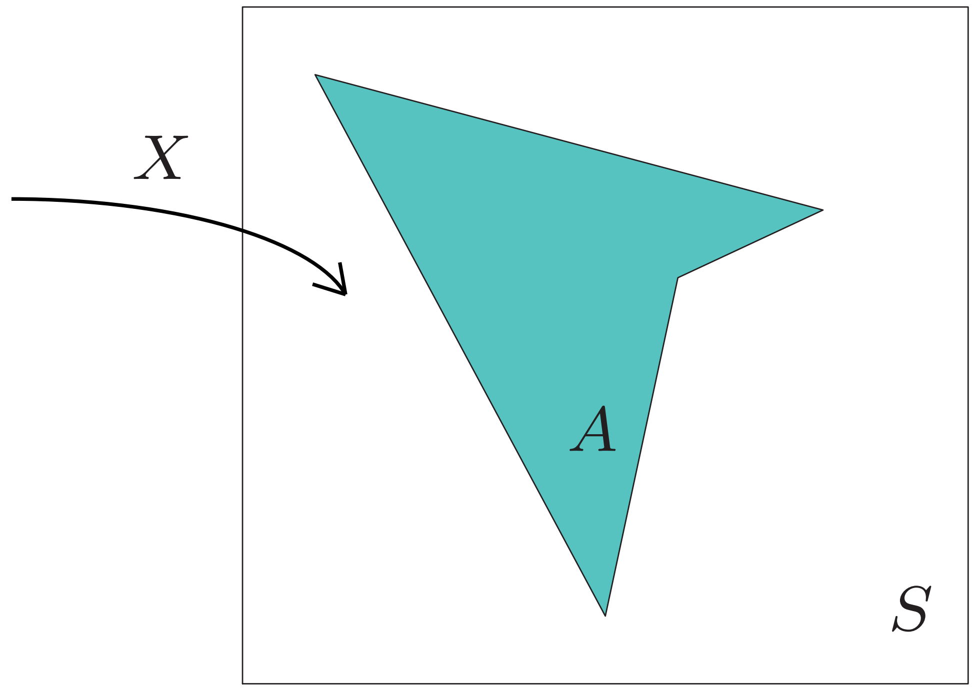
jede (messbare) Teilmenge A von S

kommt mit einer W’kt zum Zug,

die dem relativen Anteil ihres Volumens

am Volumen von S entspricht:

$$\mathbf{P}(X \in A) = \frac{\text{Volumen von } A}{\text{Volumen von } S}$$



X
A
S